5장 확률과정

정상확률과정, 자기상관함수, 부분자기상관함수

확률과정

- 확률과정(Stochastic Process): 확률법칙에 의해 생성되는 일련의 통계적 현상

- 시계열에서 다루는 대부분의 확률모형: 확률과정을 설명하는 모형

- 관측된 시계열: 확률과정이 실현된 값(관찰값)

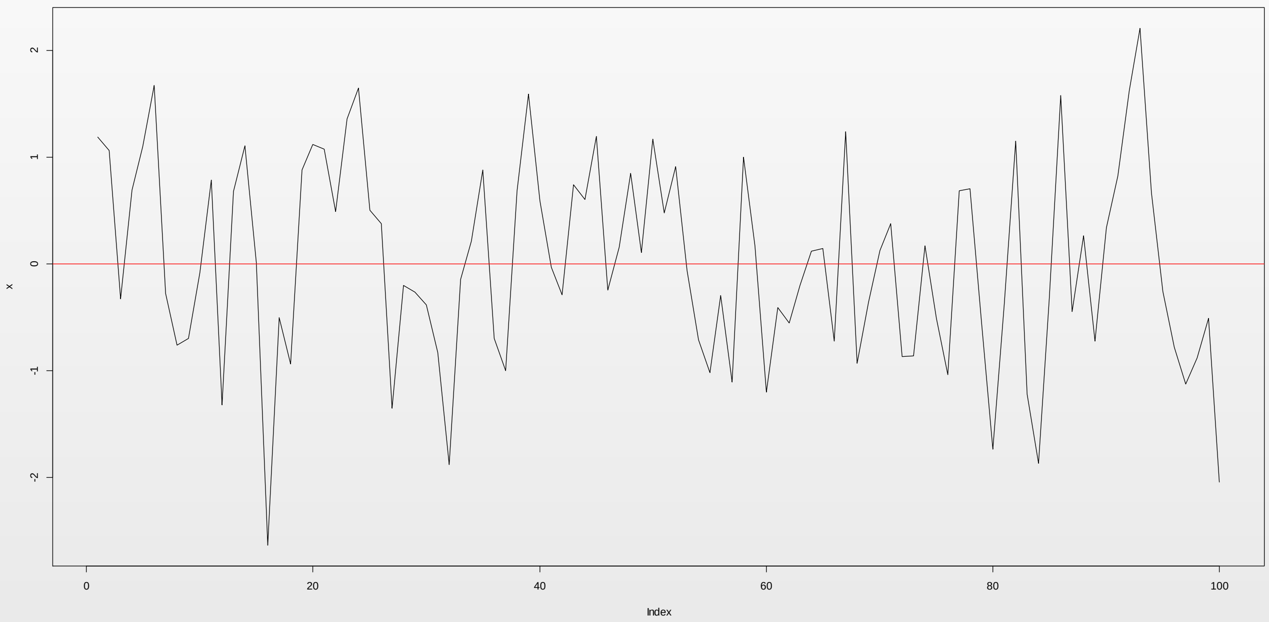
5.1 정상확률과정

- 주어진 시계열 자료 $\{Z_1, Z_2, \cdots, Z_n\}$: 특정 확률과정이 실현된 값

- 주어진 시계열 자료를 생성시킨 확률과정의 모형은 무엇인가?

- 정상성: 가능한 무수한 시계열모형 중 특정 성질을 가진 일부의 모형만을 고려 대상으로 하기 위한 개념

- 정상 시계열의 예



- 뚜렷한 추세가 없다: 평균이 같다.
- 시계열의 진폭이 일정하다: 분산이 같다.

- 강한 의미의 정상성 조건

  - 시간 축을 k만큼 이동해도 모든 n에 대하여 결합확률밀도함수가 동일

$$f(Z_{t_1}, Z_{t_2}, \cdots, Z_{t_n}) = f(Z_{t_{1+k}}, Z_{t_{2+k}}, \cdots, Z_{t_{n+k}})$$

  - 주어진 확률과정에 대한 위 조건의 만족 여부를 증명하는 것은 매우 어려움

  - 약한 의미의 정상성 조건을 주로 이용

- 약한 의미의 정상성 조건

  - 일반적 의미의 정상성 조건

  - 주어진 확률과정이 다음의 세 가지 조건을 모두 만족시키면 정상확률과정이라고 한다.

    1. $E(Z_t) = \mu$
    2. $Var(Z_t) = \sigma^2$
    3. $Cov(Z_t, Z_{t+k}) = \gamma_k$

    - 평균과 분산은 각각 상수로 시간 t과 관계없이 동일
    - 자기공분산은 시점 t와는 관계없이 시차 k만의 함수

5.2 확률과정의 예

1. 백색잡음과정
2. 확률보행과정
3. 자기회귀과정
4. 이동평균과정

1) 백색잡음과정(White noise process)

- 서로 독립이고 동일한 분포를 따르는(i.i.d.) 확률변수로 구성된 확률과정

$$Z_t = \varepsilon_t, \qquad t = 1,2,\cdots$$

ε_t는 서로 독립이고 평균이 0, 분산이 σ^2인 확률변수

- 분포가 정규분포를 따르는 경우 Gaussian white noise process라고 함

- 백색잡음과정은 정상성을 만족

- 많은 확률과정이 백색잡음과정에서 생성

2-1) 확률보행과정(Random walk process)

- $\{\varepsilon_t\}$가 백색잡음과정일 때 다음의 $\{Z_t\}$를 확률보행과정이라고 함

$$Z_t = Z_{t-1} + \varepsilon_t, \qquad Z_0 = 0, t = 1,2,\cdots$$

- 확률보행과정의 다른 표현

$$\begin{aligned} Z_1 &= \varepsilon_1 \\ Z_2 &= \varepsilon_1 + \varepsilon_2 \\ &\vdots \\ Z_t &= \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t \end{aligned}$$

- 확률보행과정의 정상성 판단

$$E(Z_t) = E\left(\sum_{i=1}^{t} \varepsilon_i\right) = 0$$

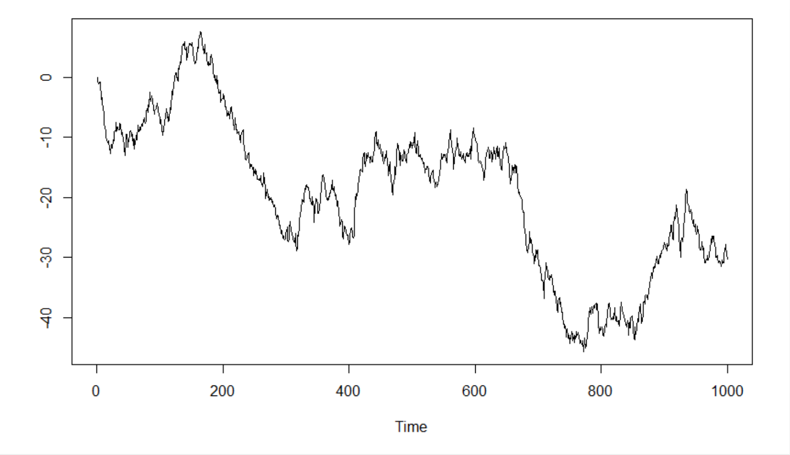
$$Var(Z_t) = \sum_{i=1}^{t} Var(\varepsilon_i) = t\,\sigma^2$$

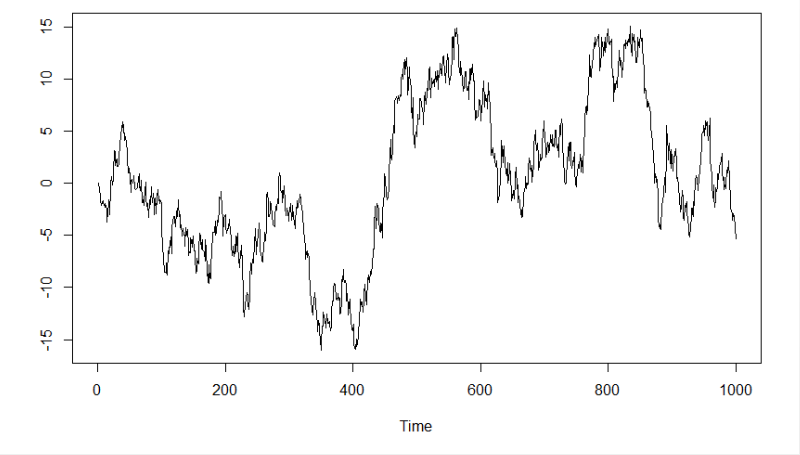
$$Cov(Z_t, Z_{t+k}) = E\left[\left(\sum_{i=1}^{t} \varepsilon_i\right)\left(\sum_{j=1}^{t+k} \varepsilon_j\right)\right] = E\left(\sum_{i=1}^{t} \varepsilon_i^2\right) = t\,\sigma^2$$

평균은 일정하지만
분산과 자기공분산은 시점 t의 함수

→ 비정상 확률과정

- 확률보행과정의 예


0
-10
-20
-30
-40
0
200
400
600
800
1000
Time


15
10
5
0
-5
-10
-15
0
200
400
600
800
1000
Time

2-2) 절편이 있는 확률보행과정(Random walk process with drift)

- 확률보행과정: 주가(stock price), 환율(exchange rate) 등의 모형에 주로 사용됨

- 경제가 호황인 경우 주가는 상승되는 추세를 보이는 것이 일반적인 상황. 확률보행과정으로는 적절한 설명이 어려움.

- 절편이 있는 확률보행과정

$$Z_t = \delta + Z_{t-1} + \varepsilon_t, \qquad Z_0 = 0, t = 1,2,\cdots$$

$$\begin{aligned} Z_1 &= \delta + \varepsilon_1 \\ Z_2 &= 2\delta + \varepsilon_1 + \varepsilon_2 \\ &\vdots \\ Z_t &= t\delta + \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t \end{aligned}$$

- 절편이 있는 확률보행과정의 정상성 확인

1. $E(Z_t) \quad = E\left(t\delta + \sum_{i=1}^{t} \varepsilon_i \right) = t\delta$

2. $Var(Z_t) = Var\left(t\delta + \sum_{i=1}^{t} \varepsilon_i \right) = t\sigma^2$

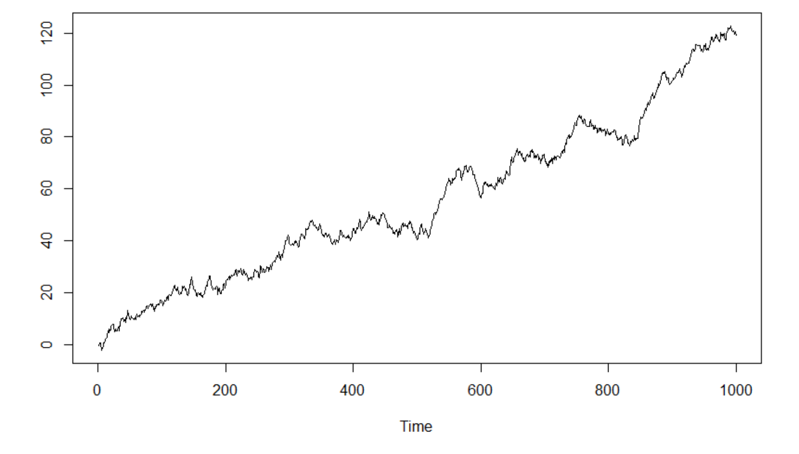
3. $Cov(Z_t, Z_{t+k}) = E(Z_t Z_{t+k}) - E(Z_t)E(Z_{t+k})$

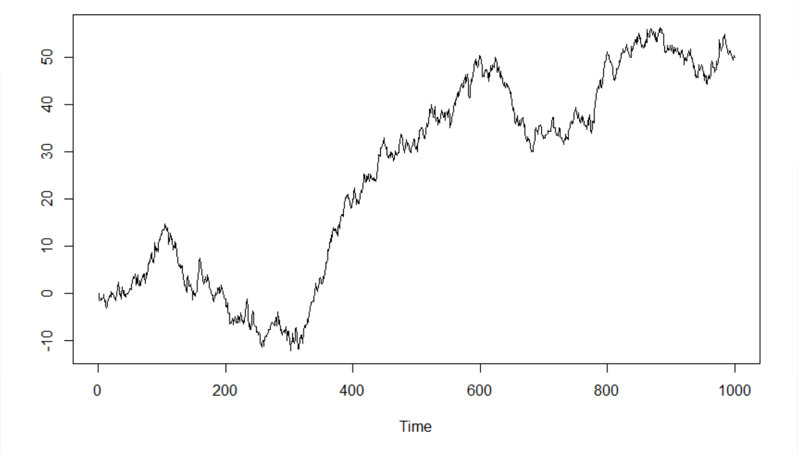
$$= E\left[\left(t\delta + \sum_{i=1}^{t} \varepsilon_i \right)\left((t+k)\delta + \sum_{j=1}^{t+k} \varepsilon_j \right) \right] - E\left(t\delta + \sum_{i=1}^{t} \varepsilon_i \right) E\left((t+k)\delta + \sum_{j=1}^{t+k} \varepsilon_j \right)$$

$$= t\sigma^2$$

비정상 시계열

- 절편이 있는 확률보행과정의 예





3) 자기회귀과정(Autoregressive process)

- 현재 시점의 자료는 과거 시점의 자료와 현재 시점의 오차에 의해 결정되는 모형

- p차 자기회귀과정: AR(p) 모형

$$Z_t - \mu = \phi_1(Z_{t-1} - \mu) + \phi_2(Z_{t-2} - \mu) + \cdots + \phi_p\left(Z_{t-p} - \mu\right) + \varepsilon_t$$

$$E(Z_t) = \mu$$

$$\varepsilon_t \ \ iid \ \ (0, \sigma^2)$$

- AR(1) 과정

  - 모형: $$Z_t - \mu = \phi(Z_{t-1} - \mu) + \varepsilon_t$$
    $$Z_t = \mu - \phi\mu + \phi Z_{t-1} + \varepsilon_t$$
    $$= \delta + \phi Z_{t-1} + \varepsilon_t$$

  만일 ϕ=1이면 AR(1)과정은 확률보행과정과 동일

  - 평균, 분산, 공분산

    $E(Z_t) = \mu$ by assumption

    $$Var(Z_t) = \frac{\sigma^2}{(1-\phi^2)}, \quad Cov(Z_t, Z_{t+k}) = \phi^k \frac{\sigma^2}{(1-\phi^2)}$$

  - 정상성 조건

    정상성 만족: 분산과 자기공분산이 유한(finite)이어야 함

    AR(1)의 정상성 조건: $|\emptyset| < 1$

4) 이동평균과정(Moving average process)

- 현 시점의 자료는 현 시점의 오차와 과거 시점의 오차들의 선형결합으로 결정되는 모형

- q차 이동평균과정: MA(q) 모형

$$Z_t - \mu = \epsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

$$E(Z_t) = \mu$$

$$\varepsilon_t \;\; iid \;\; (0, \sigma^2)$$

- MA(1) 과정

  - 모형: $Z_t - \mu = \varepsilon_t + \theta\varepsilon_{t-1}, \qquad t = 1,2,\cdots$

  - 평균, 분산, 공분산:

$$E(Z_t) = \mu$$

$$Var(Z_t) = (1+\theta^2)\sigma^2$$

$$Cov(Z_t, Z_{t+k}) = \begin{cases} \theta^2\sigma^2, & k = 1 \\ 0, & k \geq 2 \end{cases}$$

  - 정상성:

    조건 없이 항상 정상성을 만족함

5.3 자기상관함수 (Autocorrelation function: ACF)

- 시계열 자료의 특성: 현재의 상태가 과거 및 미래와 연관되어 있음

- 정상성을 만족하는 확률과정에 대하여 시간에 따른 상관 정도를 측정하는 도구가 필요함

- 자기 공분산함수(Autocovariance function)

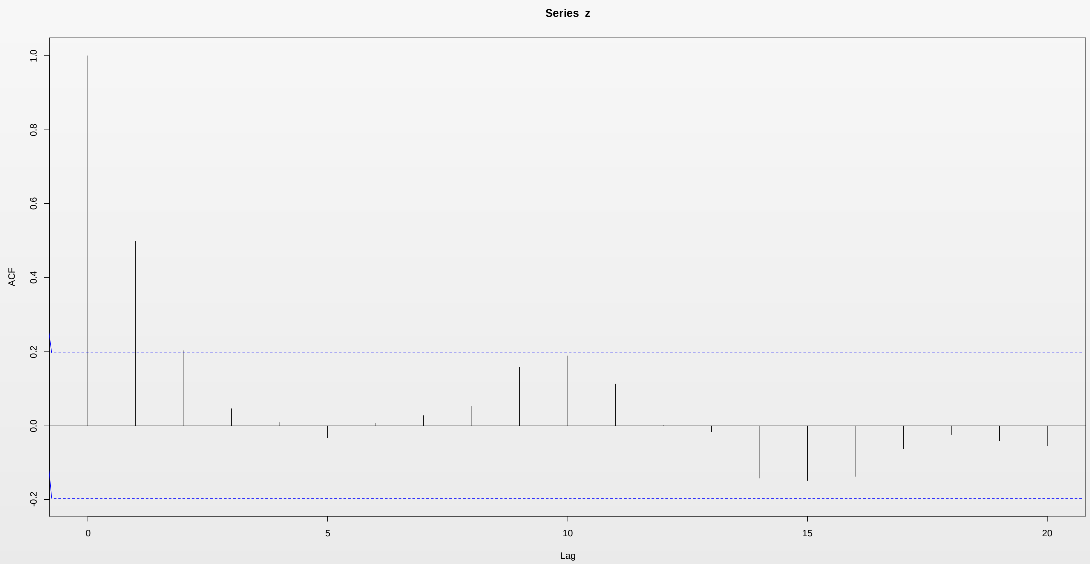
$$\gamma_k = \mathrm{Cov}(Z_t, Z_{t+k})$$

- 자기 상관함수(Autocorrelation function: ACF)

$$\rho_k = \frac{Cov(Z_t, Z_{t+k})}{\sqrt{Var(Z_t)}\sqrt{Var(Z_{t+k})}} = \frac{\gamma_k}{\gamma_0}$$

- 표본상관도표(sample correlogram): SACF

  - X축은 시차(k), Y축은 표본 ACF($\hat{\rho}_k$), k=0,1,2,...



R 함수 acf( )로 작성

  - 만일 $\rho_k = 0$이라면, 표본 ACF는 점근적으로 평균이 0, 분산 1/n인 정규분포를 따른다.

  - 이 성질을 이용한 ρ_k의 95% 신뢰구간이 점선으로 추가됨

- 표본 상관도표에 추가된 95% 신뢰구간의 이용

  - 만일 $\hat{\rho}_k$이 점선을 벗어나 있다면 귀무가설 $H_0: \rho_k = 0$을 5% 유의수준에서 기각할 수 있음

  - 이러한 방식의 검정은 다중 검정에 해당

    검정 대상 가설: $H_0: \rho_k = 0, \ k = 1,2,3, \cdots$

  - 일종오류가 증가하는 문제

  - 엄격하게 검정을 실시하는 것보다는 참고자료로 사용

- 다중 검정의 문제: Type 1 error rate의 증가
  - 3개 귀무가설 $H_0: \rho_k = 0,\ k = 1,2,3$이 모두 사실이라고 가정
  - 3개 독립된 검정을 각각 유의수준 α에서 실시
  - 3개 검정 모두에서 옳은 결정을 내릴 확률

    $P(3\text{개 검정에서 모두 } H_0 \text{ 기각 못함}) = (1-\alpha)^3 < (1-\alpha)$
  - 다중 검정의 일종 오류 확률

    $P(3\text{개 다중 검정에서의 일종 오류})$

    $= P(\text{적어도 한 번은 } H_0 \text{ 기각})$

    $= 1 - P(3\text{개 검정에서 모두 } H_0 \text{ 기각 못함}) = 1-(1-\alpha)^3 > \alpha$

예: $1-(1-0.05)^3 = 0.1426,\ \ 1-(1-0.05)^{15} = 0.5367$

5.4 부분자기상관함수 (Partial autocorrelation function: PACF)

- Z_t와 Z_{t+k}의 관련성을 판단하는데 ACF와 함께 유용하게 사용되는 통계량

- Z_t와 Z_{t+k}사이의 직접적인 관련성을 측정

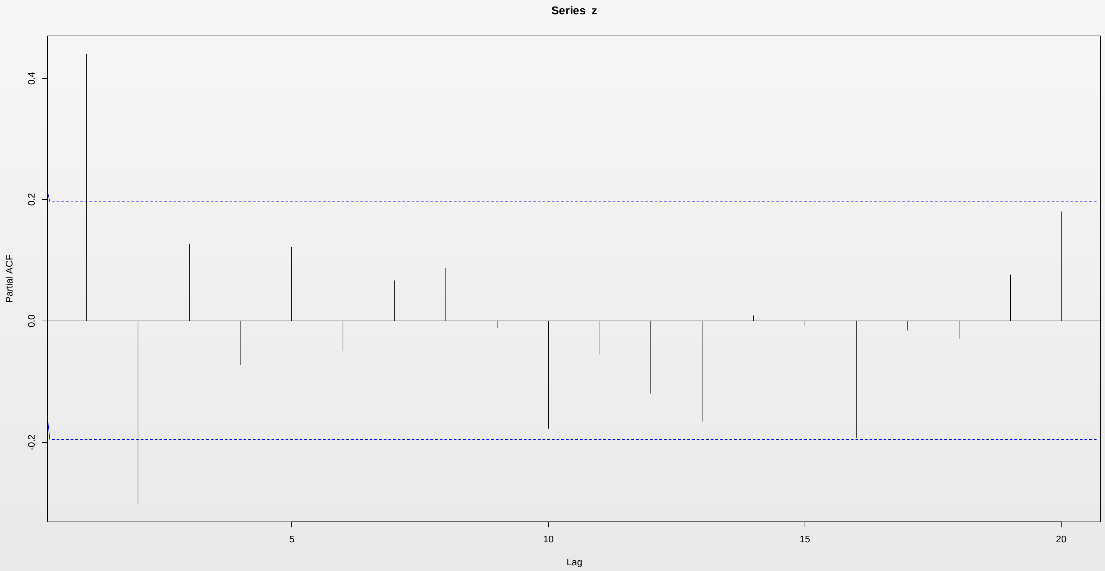
- 부분자기상관함수(PACF): $\emptyset_{kk}$

  시계열자료 $Z_t, Z_{t+1}, \ldots, Z_{t+k-1}, Z_{t+k}$가 주어졌을 때

  PACF $\emptyset_{kk}$는 Z_t와 Z_{t+k} 에서 $Z_{t+1}, \ldots, Z_{t+k-1}$의 효과를 제거한 후 상관관계를 측정

- 표본부분상관도표(sample partial correlogram): SPACF

X축은 시차(k), Y축은 표본 PACF($\hat{\phi}_{kk}$), k=1,2,…



R 함수 pacf( )로 작성

ACF의 경우와 같이 95% 신뢰구간이 함께 표시됨